

## タンパク質の立体構造からの機能予測

木下賢吾

. はじめに

センターニュースの過去の記事を見ると、どうもこの原稿は「バイオインフォマティクスシリーズ」の第3回になるようである。第1回目が2003年5月号の由良さんの記事<sup>1</sup>、第2回目が同年11月号の白井さんの記事<sup>2</sup>である。とはいっても明確にシリーズ化されているわけではなく、前二回の内容はあまり関連がない。しかし、何故か2人とも「バイオインフォマティクス」という言葉に関する緒言から始まっている。由良さん曰く「バイオインフォマティクスは情報学であると言うことをよく聞く。そういう発言を耳にするたびに、筆者は唖然とする。バイオインフォマティクスは生物学である」とおっしゃっている。一方、白井さんは、「バイオインフォマティクスは流行である」と切って捨てる一方「しかし、いまひとつしっくりこない気もする」としてバイオインフォマティクスの現状を的確に分析している。お二人とも旧知の間柄だが、おそらく期せずして同じような問いかけから話を始める必要があったのは何故だろうか？最近でこそ「ご専門は？」と聞かれたら、筆者は愛想良く「バイオインフォマティクス」ですと答えられるようになったが、1年以上前まではしっくり来なくて、できるだけ「専門は生物物理です」と答えるようにしていた。今もしっくり来ないことには変わりはないが、あまり気にならない程度に頻繁にバイオインフォマティクスという言葉が利用されるようになってきたのは事実である。しかし、なぜわざわざ「バイオインフォマティクスとは？」から始める必要があったのだろうか？

最も単純な理由は、バイオインフォマティクスという分野が新しいからだろう。先に挙げた白井さんのニュースによると、この言葉が使われるようになったのは1993年頃からだそうである。そのころは、バイオインフォマティクスの訳語としても「生物情報学」と「情報生物学」が混在していたように思う。現在は、基本的にバイオインフォマティクスの訳語としては生物情報学ということになっていて、情報生物学というのはもう少し別のニュアンスで利用される。(ちなみに、情報生物学の英語訳はinformation biologyということになっているみたいだが、この単語をgoogleで検索すると何故か日本の研究室のホームページばかりが上位に来て、欧米圏の研究室はほとんどヒットしないので、和製英語かもしれない)新しい分野であるが故に、最初にその立ち位置をはっきりとさせる必要があったということだろう。しかし、個人的にはもう一つの理由として、この分野が複合領域的な分野だからということも大きな理由だと思われる。

世界的に見ると、情報系の人たちが彼らの持つ技術を生物学に応用し、便利なツールを開発しましたというのがバイオインフォマティクス人口の多数を占めるように思える。しかし、少数派として、生物や物理化学をバックグラウンドとして計算機を利用した生物学っぽい仕事をしてい

たら、気がついたらバイオインフォマティクスの人だという分類がなされた人も存在する。筆者もその一人だと思っているし、由良さん、白井さんも（多分）同じ少数派だと思うので、似たようなしっくり来ない感を持っているのだろう。こういう一群の人たちは、正当なインフォマティクスの人たちから見ても、バイオロジーの人から見てもちょっと違うよねって感じでとらえられていると勝手に思っている節がある。半分は勝手な被害妄想だと思うが、確かに何処に行っても違和感があることは否定できない。

#### ・原理主義とオーム主義

このような違和感はどこから来るのだろうか？古今東西、異文化が出会った際にお互いを理解するまでに時間がかかり、時には衝突してしまう例は枚挙にいとまがない。バイオロジーとインフォマティクスでもしかりである。さらにここに、物理や化学のマインドを持った人たちまで絡んでくると話はややこしい。誤解を恐れずに思いっきり単純化して言えば、物理の人はさまざまな現象を単純な原理で説明しようとする。化学の人もそういう傾向はあるが、物理の人よりは少し多様性に寛容である。生物の人は、一般的な原理からははずれた現象の発見を尊しとし、とにかくさまざまな現象を枚挙することに余念がない。インフォマティクスの方は、原理原則と言うよりはテクノロジーとしての側面が強いが、マインド的には物理に近く、個別の応用例よりはそれらを抽象化した考え方としてのアルゴリズムに関心をもつ。このように程度の差はあるが、バイオインフォマティクスに関わる人たちは、原理原則を重んじる原理主義から、とにかく観測した現象を蓄積していくオーム主義的な人の集合であるととらえることができる。狂信的な前者の人は根っからの後者の人の「理解」を理解することができない。幸い極端な人はあまりいないので、真っ向から衝突することはあまりないが、お互いが大なり小なり「相手の言っていることがよく分からない」という思いを持っていると思われる。さらに、どちらかに局在してしまえばそれなりに同士がいるので違和感を感じる機会は少ないが、どちらにも完全には所属しきれていない少数派は、異なる価値観の狭間でなんとなく違和感を感じる機会が多くなる。オーム主義と原理主義のどちらが正しいとかエライとかいうことはないと思うが、生物学がオーム的な価値観と親和性が高いので、オーム主義的な仕事での成果の方が華々しい。その最たるモノがヒトゲノム解析である。

ヒトゲノム解析は2001年にほぼ完了したという宣言（ドラフト配列の公開）が公的機関の International Human Genome Sequencing Consortium (IHGSC)<sup>3</sup>及び民間のCelera社からなされて<sup>4</sup>一段落したが、正式なヒトゲノムの配列が発表されたのはごく最近のことである<sup>5</sup>。

#### ・文字列がわかっても意味が分からない

ゲノム配列が解かれて何が分かったのだろうか？ヒトゲノムに関してはまさに解かれたところであるのでまだまだこれから解析をまたなければならないが、すでに200種を超える微生物ゲノムが解かれている現状を見ると、ゲノムを読み解くことにより何が見えるのかがいま見えてくる。

ゲノムの上には、おおざっぱに言って、実際に生体内で働くタンパク質に翻訳される遺伝子の

部分と、その遺伝子の発現を制御（どういうタイミングでどれぐらいの量がタンパク質になればいいのか）する制御領域がある。生物を分子レベルで理解するためには、遺伝子産物であるタンパク質の機能を同定しなければならない。この機能同定をすべてのタンパク質に対して実験的に行うことが困難であるところに、計算による機能予測の必要性が生ずる。現在最も有用な方法は、機能既知のタンパク質との配列類似性の検索である。この方法では、配列 類縁関係 機能という相関関係を利用するため、できる限り遠い類縁関係を検出する努力がなされる。しかし、遠い類縁関係から機能を推定することには原理的に限界があることを考えると、進化的類縁関係のみに依拠しないアプローチが必要になってくる。実際配列の解析だけでは、50%程度の遺伝子の機能が推定されているだけで残りの遺伝子機能は未知のままである。

配列 機能相関を用いた機能予測の原理的な限界は、タンパク質機能の第一原理（配列 立体構造 機能）において立体構造の部分をスキップしたことによって引き起こされている。したがってその限界を超えるためには、立体構造 機能の相関に基づいた予測法を確立しなければならない。近年の立体構造情報の急速な増大及び構造ゲノム科学プロジェクトの進展<sup>6</sup>によって、第一原理的機能予測を開発する基盤が整いつつあるといえる。しかし、それにもかかわらず、立体構造情報は限定的にしか利用されていないのが現状である。

第一原理機能予測法を開発するためには、まず立体構造 機能相関の系統的な解析が必要になる。そのための課題をここに上げてみよう。第一に、機能に関する情報を整理しなければならない。例えば、伸長因子というタンパク質は、タンパク質合成の際にアミノ酸の重合をすすめるために不可欠な要素として同定されたため、生物学的にはこのような名前では呼ばれているが、化学的にはGTPの加水分解反応を触媒する。このような記述は文献の中にはあっても、例えば酵素番号のように系統だった体系からは除外されており、系統的な立体構造 機能相関の解析を大きく妨げている。第二に、機能を表現する立体構造の記述の方法を確立しなければならない。例えば、同じ機能を持ちフォールドが異なる一群のタンパク質を、機能部位周辺の原子の空間配置に着目して分類すると、スーパーファミリーに対応する分類があらわれる<sup>7</sup>。このことは、原子の空間配置それ自体では、進化的類縁関係を超越する機能の記述法にはなり得ないことを示唆している。一方で、分子表面に着目して機能部位を記述すると、そのような記述のなかには進化的類縁関係を超越した共通性が見えて来る<sup>8</sup>。これは、機能の類似性に対するひとつの必要条件であるが十分条件ではない。予測のためには、必要かつ十分な情報をもった構造の表現法を見つける必要がある。第三に、タンパク質のダイナミクスの系統的な研究の必要性である。結晶解析やNMRで解かれているタンパク質の立体構造は平均構造に過ぎず、実際はタンパク質にはゆらぎがあり、さらには機能発現に伴って構造変化をする。立体構造の表現法とも関係して、立体構造を動き得るものとして、単なる原子座標でない表現の仕方を模索することも視野に入れる必要がある。

また、一口にタンパク質の機能といっても見方によってその定義はいろいろある。例えば、いわゆるモータータンパク質はアクチンフィラメントなどと相互作用して細胞内の物質輸送や筋肉細胞の運動に関与するタンパク質であり、生物学的には、その機能はその名のとおりに「モーターとして働くタンパク」と考えられる。その一方でATPアーゼ活性を持つのでATP結合能 + 加水分解

解活性という生化学的な機能を持つとも言える。最近ではこれらを区別するために、前者を cellular function, 後者を molecular function (以下「分子機能」と呼ぶ) と呼ばれる。また、分子機能と言っても、化学反応を素過程に分けて考えるようにそれらは多くのステップに分かれており、どのような部分反応がどのように立体構造と対応しているのかはよくわかっていない。そこで、とりあえずは機能の最初のステップである結合、つまり分子認識を機能だと考えて結合部位の予測から話を進める。つまり、似た分子表面の形状と静電ポテンシャルを持つタンパク質は似た低分子リガンドを結合すると仮定する。もちろんこの仮定(どの程度似ていれば機能に対応するのか)は定量的に検証可能な仮定であり、筆者らも系統的にこの仮定の検証を進めているが、ここでは紙面の都合上話をシンプルにするためにこの仮定を前提として話を進める。なお、ここでは結合能のみを問題にするので、酵素タンパク質以外でも、タンパク質に結合する低分子を区別せずにリガンドと呼ぶことにする。

#### ・ 構造からの機能予測

以上のように原理的には難しい問題はいろいろあるが、リガンド結合能だけに話を絞って、構造からの機能予測の問題を考えてみたい。

リガンドとタンパク質の相互作用を予測する方法はいくつか提案されている。これらは大別すると2つに分けられる。(1) 評価関数を用いた方法,(2) 類似性検索を用いた方法。

(1) 評価関数を用いた方法は、その評価関数が物理化学的なものかデータベースから抽出された経験的ポテンシャルであるかという違いはあるが、基本的には分子間の相互作用のもっともらしさを評価する関数を最適化することにより結合様式の予測を行う。この方法の利点は、その評価関数が一般的な特徴を十分に反映していれば、データベース中に存在しない相互作用様式があっても正しく予測を行うことができる可能性があるという点である。そのため、できるだけ物理化学的にもっともらしい、つまり、より一般性があると思える相互作用エネルギー関数を用いるが、それぞれのエネルギー関数をもっともらしくなるにしたがって、タンパク質の立体構造情報だけからは扱いたい水やタンパク質の柔軟性の効果を取り扱う必要が出てくる。しかし、水やタンパク質の構造揺らぎをあらわに扱うことは、非常に計算時間がかかるため、実用的に数多くのタンパク質に応用することは当面は不可能であると考えられる。一方、水の扱いが不十分だとよい結果が得られないため、さまざまな効果を平均的に取り込んだ、データベース依存の経験的な評価関数を用いることがよくなされる。しかし、「さまざまな効果を平均的に」取り込むことは、相互作用の様式に共通の特徴が多くある場合には有効に機能するが、相互作用の様式が多様な場合には、平均値に意味がないので、うまく機能しない。実際、タンパク質・DNA相互作用のように、その相互作用様式の支配的な因子が、静電ポテンシャルであるというように共通性が高い場合には、さまざまなタンパク質DNA複合体の結合部位によく見られる相互作用様式を経験的ポテンシャルの形式に直して利用する、単純なアプローチでも80%を超える正答率が得られるが、タンパク質・DNAの相互作用でも非常に特徴的な相互作用をしている場合にはこのアプローチではうまく予測できない<sup>9</sup>。また、リガンド複合体予測に同様の単純な経験的ポテンシャルを利用した

だけでは高い正答率が得られない (unpublished results)。他の研究グループでも経験的ポテンシャルによるタンパク質間相互作用の予測がうまくいっていないのは、ここに原因があると思っている。つまり、多様な相互作用様式を持つタンパク質間相互作用には平均値を扱い、個別の相互作用を無視してしまう経験的ポテンシャルは機能しないのではないだろうか。

(2) これに対して類似性検索では、すでに複合体で構造が解かれているタンパク質の結合部位をデータベース化し、そのデータベースに対する類似性検索を行うことにより結合部位の予測を試みる。類似性検索によるアプローチでは、既知の複合体のデータベースに依拠するので、データベースにない相互作用様式はうまく予測できないという問題があるが、多様な相互作用様式をまるごとデータベース化してしまえば、経験的ポテンシャルで問題になる、まれな相互作用様式を評価できない問題は回避可能である。通常は、検索に時間がかかるという問題があるため、すべてをデータベースに入れて検索をおこなうというアプローチは現実的でないが、筆者らはグラフ理論に基づいた高速な類似性検索法を開発することでこの問題を解決した<sup>10</sup>。そこでつぎに解決すべき問題は、データベースにない相互作用様式をどのように扱うかである。しかし、近年立体構造解析の進展はめざましく、たんぱく3000のような構造ゲノム科学プロジェクトの進展により、複合体で構造が解かれるタンパク質は急激に増えておりこの問題は徐々に解決に向かっていく。さらに、申請者が低分子であきらかにしたように、結合部位全体をひとまとまりにしてとらえるのではなく、部分構造に分けてデータベース化することにより、全体では今まで見られなかったような結合様式も部分の組み合わせとして記述できれば、部分構造データベースに対する類似性検索で、全体としては新規な結合様式を持つ結合部位も予測できるようになると予想される。

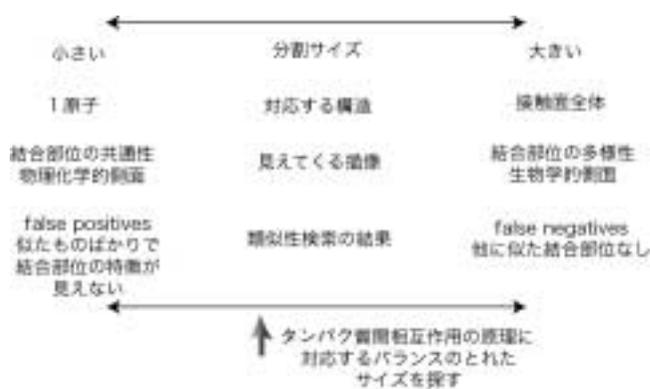


図1 分割サイズと共通性・多様性のバランス関係

タンパク質間相互作用部位を部分構造に分けてデータベース化する課程で重要になるのは、共通性と個別性のバランスである。このバランスは部分構造をどれぐらいの大きさに分けるかで決まる(図1)。小さな部分に分ければ分けるほど、物理化学的な相互作用と対応する相互作用様式が出てくるか、その一方で大量に似

ている部分が見いだされ、本当に重要な類似性が他の重要でない類似性の中に埋もれてしまう可能性があるが、それでは経験的ポテンシャルを利用したときと同じことになってしまう。だからといって、ほとんど部分に分けなければ、すべての相互作用様式が別物になり、類似性検索の弱点である、データベース中に存在しない相互作用様式は予測できないという問題の解決にならない。そこで、結合部位の部分構造の比較分類と試行錯誤により、共通性と個別性のバランスを探ることになるが、この課程を通じて、タンパク質間相互作用の多様性と共通性をあきらかにした

いと考えている。

以上のことを考え合わせると、たんぱく3000のような構造ゲノム科学プロジェクトの進展により立体構造データベースがますます充実することを念頭におくと、類似性検索によるアプローチでリガンド結合部位の予測をおこなうことはうまくいく可能性が高いと考えられる。

#### . eF-site database & PDBjViewer

我々は以上のような方針でリガンド結合部位に着目した分子表面の静電ポテンシャルのデータベースとしてeF-site/Binding-Siteを構築した (<http://eF-site.protein.osaka-u.ac.jp/eF-site>)。現在このデータベースには2003年はじめまでにX線結晶構造解析により立体構造が解析されたタンパク質のうち、2.5Åより良い分解能をもち、低分子と結合状態が登録されているタンパク質のほとんどすべて(金属, SO<sub>4</sub>, PO<sub>4</sub>を除く22,747結合部位)がすでに登録済みである<sup>11</sup>。このデータベースでは、Java3Dを使ったviewerも開発することにより、タンパク質のリボンモデルと分子表面モデルを同時に対話的に動かして見ることが可能になっている。このviewerはPDBj ([www.pdbj.org](http://www.pdbj.org))の活動の一環として開発し、ソースコードを無償で公開している。

#### VI. 類似性検索法と類似度の規格化

類似性検索法の詳細に関しては参考文献<sup>8, 11</sup>を見て頂くとして、ここでは応用例を理解するのに必要な最小限度の説明を簡単に述べておく。

分子表面はそもそも連続的であるが、実際には計算機で扱いやすいように3角ポリゴンで表現している。このポリゴンの各頂点での曲率及び静電ポテンシャルを考えて、これらが似た値をもつ頂点を対応付け、その対応付けにしたがって重ね合わせを行ったときに *root mean square deviation* (*r.m.s.d*) が1.0Å以下であるという条件の下でできるだけ多くの頂点对応づけられる対応付けの方法を探す。結果として2つの表面を比較したときに出てくる類似度は、対応する頂点の数で測ることになる。1つの問い合わせ構造のタンパク質をデータベース全体に対して検索を行うと、登録済みの22,747個の結合部位との類似度(対応する頂点の数)を得ることになる。

この類似度で良く似ていると判断される結合部位を探せば良いのだが、この類似度をそのまま利用するには少々問題がある。一つは、非常に一般的な問題として、さまざまなタンパク質を問い合わせ構造にした時にその結果を同じ基準で評価する時の問題である。頂点数をそのまま利用していると、大きなタンパク質を問い合わせ構造にした時には平均的に多くの頂点对応するので、多くの結合部位が類似していることになってしまうが、これは望ましい結果ではない。そこで通常よく行われるのは、問い合わせ構造として使ったタンパク質ごとに、得られた類似度の平均と分散を利用して規格化する(この規格化された類似度はZ-Scoreと呼ばれる)。もう一つの問題は、ほぼ同様の問題が検索するデータベースの結合部位の大きさでも生じる。つまり大きな結合部位(多くの頂点を含む部分表面)では平均的により多くの頂点对応する可能性があるため、より大きな類似度を示す。つまりより大きな結合部位が大きなZ-scoreを持つ傾向がある。もちろんこれは望ましいことではない。そこで結合部位に関しても同様の規格化を行う必要がある

が、さまざまな問い合わせ構造に関してそれぞれの結合部位の部分表面が平均的にどの程度の数の対応する頂点があるか、またそれがどのような分散を持つかを見積もるのは計算量の観点から現実的ではない。そこで我々は、対応する頂点が結合部位の部分表面に占める割合（coverage）をZ-scoreと同時に利用することにより結果を解釈することとした（例としては後で述べる図2）。このような規格化を利用すると、Z-score、coverageともに大きな値を持つ、つまり図2では右上の方に来る結合部位がより有意だと考えられる。しかし、どの程度右上に来ていけば統計的に有意であるかないかの閾値はいまのところははっきりできていない。現在はいくつかの例をもとに図2にある階段状の関数を設定しているが、これは暫定的な閾値線であり、今後応用例を積み重ねることにより、既知の例とうまく対応する閾値に徐々に改善していく必要がある。ここで単純な直線ではなく階段状の関数を採用したのは、左下の似ていない結合部位の分布の縁の形を模したからである。

## Ⅶ．応用例

以上のような方法を実際にいくつかの機能未知タンパク質に応用し、良好な結果を得ることができた。ここではそのなかの一つとして、メタン生成古細菌（*M. Jannaschii*）由来の機能未知タンパク質 MJ0226を取りあげる。

MJ0226（PDB: 1b78）は、メタン産生古細菌のゲノム決定に伴い見いだされたタンパク質で当初は機能未知として分類されていた。Hwang<sup>12</sup>らはこのタンパク質の結晶構造解析を行うことにより、このタンパク質が結晶構造中ではホモダイマーで存在し、そのそれぞれのサブユニットが全体としては新規フォールドでありながら、部分的にモノヌクレオチド結合タンパク質と類似することを見いだした。そこで、数種類のモノヌクレオチドに対する加水分解能を測定したところ、XTP（xanthosine-5'-triphosphate）やITP（inosine triphosphate）の加水分解活性を見いだすことができた。そこで、ATPアナログ（phosphoaminophosphonic acid-adenylate ester）との結晶構造解析を行い、複合体の結晶構造を得た（PDB: 2mjp）。このような経緯の結果、MJ0226では、ATPアナログ結合状態と非結合状態の両方の結晶構造が利用可能である。そこで我々は、非結合状態の構造を問い合わせ構造として利用し、結合状態を答え合わせに利用することにより、我々の開発した表面の類似性検索による機能推定法がこのタンパク質のモノヌクレオチド結合部位を予測できるかどうかを検証した。

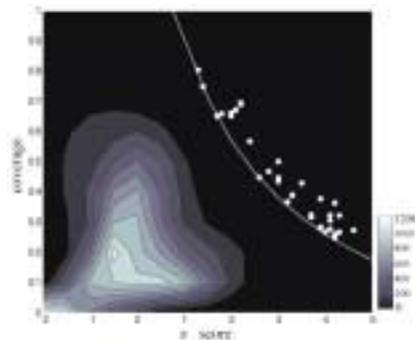


図2 MJ0226の類似性検索に関する Z-score vs. coverage plot.

図2に示すように、類似性検索の結果、現在の閾値線で似ていると判断される結合部位（白丸）は50個見つかった。これら50個の予測結合部位は結合する分子の種類、位置ともさまざまであるが、結合すると予測された分子の重心位置で分類すると16箇所に分かれることがわかった。言い換えると、我々の類似性検索の結果、MJ0226の非結合状態の構造上には16箇所のさまざまなリガンドの結合部位が予測されたことになる。

これら結合部位のうち主なものを図3に示す。この絵は、結合部位が似ていると検出できた複合体の結合部位を問い合わせ構造（今はMJ0226）に重ね合わせて、結合部位にある低分子を問い合わせ構造の分子表面上に表示したものである。今回の例で検証したかったモノヌクレオチド結合部位が、図3左の絵の左の部分にATPアナログの結合部位として予測されていることを確認できた。

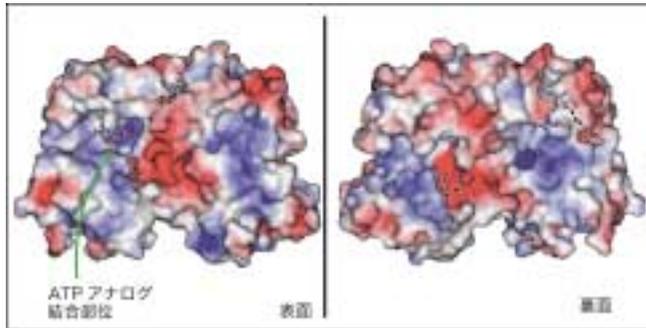


図3 MJ0226上に検出されるさまざまな結合部位。ATPアナログ欠都合部位以外にもいくつかのリガンドの仮想的結合部位が検出できる。

が、フォールドの類似性は必ずしも機能の類似性を意味するわけではない。近年の理解では、フォールドが似ていることはアミノ酸配列の類似性では検出できないほど遠い類縁関係を検出していることに対応し、場合によっては機能が保存しているために、フォールドが似ていれば機能が似ているということが観測されていると考える。つまり、構造機能相関という観点ではフォールドの類似性は機能の類似性の直接的要因ではない。しかし、現在の原子配置の類似性検索法などでは、異なるフォールド間での機能部位の局所構造の類似性を検出できることは非常にまれであり、通常似たフォールドが見いだされない場合はそれ以上の機能解析をあきらめることが多い。こういう観点からいって、今回異なるフォールド間でモノヌクレオチドの結合部位に類似性を見いだせる方法を開発できたことは、フォールドの類似性といった間接的な構造機能相関に頼らずに、構造から機能を予測する方法を開発する上で重要なハードルの一つ越えることができたと考えている。

これ以外の結合部位に関しては、今のところ実験的なサポートがないので多くを語ることはできない。しかし、モノヌクレオチド結合部位の類似性とくらべても、遜色ない類似性で予測されている部位が多いことを前向きに解釈することも可能ではないだろうか。例えば、このタンパク質ではモノヌクレオチドの加水分解能が言われているが、モノヌクレオチドの結合部位のすぐ近くにdUMPの結合部位があることから、dUMPキナーゼの可能性も考えられないだろうか？また同様に、そのほかの部位も何か別の機能に使われている可能性は考えられないだろうか。一般的に、タンパク質の機能を考える際には実験で測定された機能だけが評価され、測定をしていない他の機能があるかないかは問題にもならない。しかし実際には、タンパク質は非常に多くの機能を周りの環境に応じて使い分けているという可能性を考えるのは考えすぎだろうか？

いずれにせよ、フォールドを越えた類似性から多くの機能部位を予見できるようになってきたという状況を考えると、今後は、タンパク質の構造機能相関だけでなく、タンパク質の細胞内局

ところで、ここで強調しておきたいのだが、今回問い合わせ構造と類似性が検出された結合部位はプテロイルポリグルタミン酸合成酵素（PDB: 1jbv）やピルビン酸キナーゼ（PDB: 1a49）など、いずれもMJ0226とはフォールドが異なるタンパク質である。通常、機能予測に構造情報を利用する際には、フォールドレベルの類似性検索が行われる

在の予測（あるいは実験的に決まった情報）などと合わせて、周りにどういう分子がどの程度の数存在するのかといった情報も考えて行く必要が出てくるだろう。また、構造情報を利用する際に常に問題になる構造変化が大きなタンパク質への対応及び、その極端なケースとして、リガンド非結合状態では決まった形をとっていないループ（PDB中には座標が存在しない）の結合状態での構造予測などが重要な課題になってくると思われる。あとは機能予測とは少し別の話になるが、タンパク質がどの程度の多機能性を持つのかに関して、細胞内の代謝中間産物を網羅的に決定するメタボローム解析などのデータが出てくるとおもしろいと思っている。

#### 参考文献

- 1 由良敬, 2003年, バイオインフォマティクスが追い求めるもの, 名大情報連携基盤センターニュース第3号136 - 143
- 2 白井剛, 2003年, バイオインフォマティクスのよしなしごと, 名大情報連携基盤センターニュース第5号411 - 418
- 3 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860-921
- 4 Venter, J. C. et al. (2001) The sequence of the human genome. *Science* 291, 1304-51
- 5 International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431, 931-45
- 6 Teichmann, S. A., Chothia, C. & Gerstein, M. (1999) Advances in structural genomics. *Curr Opin Struct Biol* 9, 390-9
- 7 Kinoshita, K., Sadanami, K., Kidera, A. & Go, N. (1999) Structural motif of phosphate-binding site common to various protein superfamilies: all-against-all structural comparison of protein-monomonucleotide complexes. *Protein Eng* 12, 11-14
- 8 Kinoshita, K. & Nakamura, H. (2003) Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. *Protein Sci* 12, 1589-1595
- 9 Tsuchiya, Y., Kinoshita, K. & Nakamura, H. (2004) Structure-based prediction of DNA-binding sites on proteins using the empirical preference of electrostatic potential and the shape of molecular surfaces. *Proteins* 55, 885-894
- 10 Kinoshita, K., Furui, J. & Nakamura, H. (2002) Identification of protein functions from a molecular surface database, eF-site. *J Struct Funct Genomics* 2, 9-22
- 11 Kinoshita, K. & Nakamura, H. (2004) eF-site and PDBjViewer: database and viewer for protein functional sites. *Bioinformatics* 20, 1329-1330
- 12 Hwang, K. Y., Chung, J. H., Kim, S. H., Han, Y. S. & Cho, Y. (1999) Structure-based identification of a novel NTPase from *Methanococcus jannaschii*. *Nat Struct Biol* 6, 691-6

(きのした けんご : 東京大学医科学研究所)